Procesamiento del Lenguaje Natural, Revista nº 65, septiembre de 2020, pp. 109-112 recibido 01-04-2020 revisado 24-05-2020 aceptado 24-05-2020

Transcripción, indexación y análisis automático de declaraciones judiciales a partir de representaciones fonéticas y técnicas de lingüística forense

Transcription, indexing and automatic analysis of judicial declarations from phonetic representations and techniques of forensic linguistics

Pedro José Vivancos-Vicente[1], **José Antonio García-Díaz**[2], **Ángela Almela**[3], **Fernando Molina**[1], **Juan Salvador Castejón-Garrido**[1], **Rafael Valencia-García**[2]
[1]Vócali Sistemas Inteligentes S.L.
[2]Facultad de Informática, Universidad de Murcia
[3]Facultad de Letras, Universidad de Murcia
{pedro.vivancos, fernando.molina, juans.castejon}@vocali.net
{joseantonio.garcia8, angelalm, valencia}@um.es

Resumen: Recientes avances tecnológicos han permitido mejorar los procesos judiciales para la búsqueda de información en los expedientes judiciales asociados a un caso. Sin embargo, cuando técnicos y peritos deben revisar pruebas almacenadas en vídeos y fragmentos de audio, se ven obligados a realizar una búsqueda manual en el documento multimedia para localizar la parte que desean revisar, lo cual es una tarea tediosa y que consume bastante tiempo. Para poder facilitar el desempeño de los técnicos, el presente proyecto consiste en un sistema que permite la transcripción e indexación automática de contenido multimedia basado en tecnologías de deep-learning en entornos de ruido y con múltiples interlocutores, así como la posibilidad de realizar análisis de lingüística forense sobre los datos para ayudar a los peritos a analizar los testimonios de modo que se aporten evidencias sobre la veracidad del mismo.
Palabras clave: Reconocimiento de voz, word spotting, lingüística forense

Abstract: Recent technological advances have made it possible to improve the search for information in the judicial files of the Ministry of Justice associated with a trial. However, when judicial experts examine evidence in multimedia files, such as videos or audio fragments, they must manually search the document to locate the fragment at issue, which is a tedious and time-consuming task. In order to ease this task, we propose a system that allows automatic transcription and indexing of multimedia content based on deep-learning technologies in noise environments and with multiple speakers, as well as the possibility of applying forensic linguistics techniques to enable the analysis of witness statements so that evidence on its veracity is provided.
Keywords: Speech recognition, word spotting, forensic linguistics

1 Introducción

Los gobiernos de los países avanzados están llevando a cabo una serie de medidas de mejora y renovación tecnológica de los ministerios de justicia para mejorar los procesos judiciales y así reducir la posibilidad de errores, evitar la obsolescencia de equipos y mejorar la eficiencia en el uso de sus activos (Ballesteros, 2011). Algunas de estas medidas están enfocadas a mejorar la labor de los funcionarios de las administraciones públicas con el perfeccionamiento de sus servicios de comunicaciones internas, modernizar los sistemas informáticos, adaptarse al teletrabajo e incluir nuevas funcionalidades en sus sistemas

ISSN 1135-5948. DOI 10.26342/2020-65-15

internos.

Más concretamente, existen sistemas tecnológicos capaces de digitalizar en vídeo las vistas que tienen lugar en las salas de Justicia, como las descritas en (Zalman, Rubino, y Smith, 2019). Además, disponen de herramientas que permiten la búsqueda de información, tanto en el texto como en los metadatos y en los expedientes judiciales asociados a un caso. Sin embargo, cuando técnicos y peritos judiciales deben revisar pruebas almacenadas en documentos multimedia (como vídeos o fragmentos de audio), se ven obligados a realizar, o bien un visionado completo del vídeo, o bien probar su visualización desde cierto momento para localizar fragmentos que se consideren relevantes para el caso. Esta tarea es tediosa y consume una cantidad de tiempo considerable.

Para poder facilitar el desempeño de los técnicos es necesario poder clasificar y comprender el contenido de los recursos multimedia de manera eficiente. Los últimos avances en nuevas tecnologías permiten salvar estas dificultades con los consecuentes beneficios que ofrece disponer de una indexación eficiente del contenido multimedia. Con ello, se podrían reducir los ya dilatados procedimientos judiciales, suponiendo un ahorro a largo plazo y un beneficio para la ciudadanía.

Por otro lado, la lingüística forense, que es la rama de la lingüística aplicada que se encarga de estudiar los diversos puntos de encuentro entre lenguaje y derecho, ha demostrado su utilidad para aportar evidencias lingüísticas en los procesos judiciales. Esta disciplina se está implantando cada vez más en España y está mucho más extendida en países anglosajones, y pretende mostrar un conjunto de características lingüísticas que sea capaz de servir de apoyo en análisis del contenido de las declaraciones para determinar si ciertos testimonios o declaraciones tienen garantías de ser veraces.

El objetivo de este proyecto es el desarrollo de un sistema de transcripción, indexación y análisis automático de declaraciones judiciales a partir de representaciones fonéticas y técnicas de lingüística forense que permitan ayudar a los técnicos y peritos en el desempeño de su tarea, reduciendo así los ya dilatados procesos judiciales y mejorando la eficacia del sistema. El presente documento está estructurado de la siguiente manera. La sección 2 describe la arquitectura funcional del sistema, así como cada uno de los módulos que la componen mientras que la sección 3 describe su estado actual y nuevas vías de actuación.

2 Arquitectura del sistema

Este sistema está formado por tres módulos principales: 1) Sistema de transcripción de conversaciones de declaraciones judiciales basado en Deep-Learning (ver Sección 2.1), 2) Sistema de indexación y búsqueda de vídeos (ver Sección 2.2), y 3) Sistema de procesamiento de lingüística forense basado en características psicolingüísticas y fenómenos de duda en el discurso (ver Sección 2.3). En pocas palabras, el funcionamiento del sistema es el siguiente: A partir de un conjunto de vídeos y recursos multimedia introducidos de manera manual, el sistema transcribe automáticamente su contenido a ficheros WebVTT; en segundo lugar, estos ficheros se utilizan para construir un sistema de indexación para poder buscar sobre el vídeo a partir de citas textuales palabras relacionadas o bien a través de búsqueda fonética. Al final se provee a los peritos con una interfaz web donde pueden buscar sobre el fragmento exacto del vídeo a la vez que se genera un informe de lingüística forense que ayude a determinar la veracidad o duda de un determinado testimonio.

A continuación, se describen brevemente estos subsistemas.

2.1 Sistema de transcripción de conversaciones de declaraciones judiciales basado en Deep-Learning

El sistema de transcripción es capaz de aplicar un proceso de transcripción automática de los vídeos basada en tecnologías de Deep-Learning, mejorando su desempeño en situaciones de entornos con ruido de ambiente (Hannun et al., 2014), con varias personas hablando al unísono (Snyder et al., 2019) y sin la necesidad de un entrenamiento específico por parte de los interlocutores.

El resultado final de este primer sistema son ficheros en formato WebVTT (Pfeiffer y Hickson, 2013), que es un formato para mostrar pistas de texto cronometrado que se utilizan en sistemas de películas, vídeos en streaming, etc. Este formato permite incluir metadatos para añadir información asociada a los datos. De esta forma, se enriquecen las transcripciones realizadas añadiendo información

Marcador	**Ejemplos**
Vaguedad o exactitud	Cantidades inexactas, fechas, nombres, lugares, ...
Incertidumbre o certeza	Expresiones como *a menudo*, *hasta donde yo sé*, ...
Distanciamiento del hecho	Uso de pronombres personales en tercera persona, ...
Minimización	Expresiones como *lo que pasó*, *el incidente*, ...
Refuerzo de credibilidad	Adverbios como *honestamente*, *sinceramente*, ...
Generalizaciones	Referencias a colectivos o el uso de determinantes indefinidos, ...
Comunicación evitativa	Eufemismos, evasivas, interrupciones, ...
Actitud cooperativa	Expresiones como *tienes razón*, *es cierto*, *así es*
Egocentrismo	Uso de pronombres en primer personal del singular, ...
Repetición	Palabras duplicadas o expresiones como *ambos dos*
Exceso de lenguaje culto	Cultismos, latinismos, verbos en subjuntivo, ...

Tabla 1: Marcadores de discurso falaz

Marcadores	**Ejemplos**
Tiempo y espacio	países, capitales, lugares, fechas, rangos de tiempo, ...
Simplicidad	Uso de verbos en indicativo simple, frases cortas, ...
Experiencias subjetivas	Fenómenos percibidos a través de los sentidos
Orden	Expresiones como *de una parte*, *por un lado*, *para terminar*, ...
Detalles	Nombres propios, lugares, eventos, empresas, profesiones, ...
Naturalidad	Respuestas cortas

Tabla 2: Marcadores de discurso veraz

sobre el contexto situacional, así como qué hablante produjo la frase registrada.

2.2 Sistema de indexación y búsqueda de vídeos

El sistema de indexación y búsqueda de vídeos es el encargado de indexar y catalogar cada uno de los distintos vídeos en repositorios. Este sistema permite realizar búsquedas precisas de vídeos a partir de palabras clave, fonemas y frases parecidas e identificar y cargar el vídeo en el fragmento preciso. Para ello, este sistema es capaz de buscar en distintos tipos de formato multimedia (formatos AVI, MP4) y buscar entre la información transcrita a partir del sistema anterior y además realizar búsquedas directamente en los fragmentos de audio a través del modelado de fonemas.

2.3 Sistema de procesamiento de lingüística forense basado en características psicolingüísticas y fenómenos de duda en el discurso

El sistema de procesamiento de lingüística forense basado en características psicolingüísticas y fenómenos de duda en el discurso se encarga del análisis y extracción de características lingüísticamente relevantes relacionadas con análisis de veracidad del testimonio. Para ello, se ha desarrollado un sistema de análisis lingüístico diseñado en español basado en (Salas-Zárate et al., 2017) para evaluar distintas características para determinar si un determinado relato es veraz o, si por el contrario, es engañoso, como el estudio presentado en (Almela, Valencia-García, y Cantos, 2012). Para ello, el sistema analiza el discurso y ofrece al perito un informe que contiene fenómenos lingüísticos determinados que permiten ver si se minimizan los hechos, si hay negaciones excesivas, una actitud colaborativa, lapsus linguae, o bien si el discurso es coherente, consistente, simple, equilibrado y natural. Esta información se muestra a los peritos a través de una interfaz intuitiva que resalta dentro del texto cuáles son las palabras y frases clave identificadas por el análisis.

En la Tabla 1 y en la Tabla 2 se muestran algunos de los marcadores usados para identificar el relato falaz y veraz así como algunos ejemplos de cada marcador. Para poder extraer información del texto en base a estos marcadores se han empleado reconocedores de entidades, corpus anotados como los analizados en (Jiménez-Zafra et al., 2020) y el uso lexicones.

3 Trabajo futuro

Actualmente nos encontramos en la última anualidad del proyecto. Se está terminando el desarrollo de cada uno de los subsistemas y realizando distintas evaluaciones de rendimiento y fiabilidad.

Las tecnologías que hasta el momento nos han dado mejor resultado en la transcripción de conversaciones se basan en modelos DNN-HMM (modelos ocultos de Markov con redes neuronales profundas) (Ravanelli y Omologo, 2017). Los tipos de redes neuronales más efectivas son aquellas que, por tener en cuenta elementos pasados, se ajustan mejor a la clasificación de series temporales, como son Time Delay Neural Networks (TDNN) (Sun et al., 2017) y Long-Short Time Memory (LSTM) (Zhang et al., 2016). La indexación y búsqueda de vídeos se basa en tecnologías de Keyword Spotting, que trata de identificar una serie de palabras clave en señales acústicas. Además, se están estudiando otras variables de lingüística forense con el fin de tener en cuenta los metadatos de la conversación para medir la latencia, el tono y la velocidad del hablante.

Por último, se realizará la integración de los módulos en un prototipo final de la plataforma global. En este sentido, se planificarán distintas pruebas de campo para comprobar la calidad del sistema completo.

Agradecimientos

Este proyecto ha sido financiado por el Instituto de Fomento de la Región de Murcia con fondos FEDER dentro del proyecto con referencia 2018.08.ID+I.0025

Bibliografía

Almela, Á., R. Valencia-García, y P. Cantos. 2012. Detectando la mentira en lenguaje escrito. *Procesamiento del lenguaje natural*, 48:65–72.

Ballesteros, M. C. R. 2011. La necesaria modernización de la justicia: especial referencia al plan estratégico 2009-2012. *Anuario jurídico y económico escurialense*, (44):173–186.

Hannun, A., C. Case, J. Casper, B. Catanzaro, G. Diamos, E. Elsen, R. Prenger, S. Satheesh, S. Sengupta, A. Coates, y others. 2014. Deep speech: Scaling up end-to-end speech recognition. *arXiv preprint arXiv:1412.5567*.

Jiménez-Zafra, S. M., R. Morante, M. Teresa Martín-Valdivia, y L. A. Ureña-López. 2020. Corpora annotated with negation: An overview. *Computational Linguistics*, 46(1):1–52.

Pfeiffer, S. y I. Hickson. 2013. Webvtt: The web video text tracks format. *Draft Community Group Specification, W3C*.

Ravanelli, M. y M. Omologo. 2017. Contaminated speech training methods for robust dnn-hmm distant speech recognition. *arXiv preprint arXiv:1710.03538*.

Salas-Zárate, M. P., M. A. Paredes-Valverde, M. Á. Rodríguez-García, R. Valencia-García, y G. Alor-Hernández. 2017. Automatic detection of satire in twitter: A psycholinguistic-based approach. *Knowl. Based Syst.*, 128:20–33.

Snyder, D., D. Garcia-Romero, G. Sell, A. McCree, D. Povey, y S. Khudanpur. 2019. Speaker recognition for multi-speaker conversations using x-vectors. En *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, páginas 5796–5800. IEEE.

Sun, M., D. Snyder, Y. Gao, V. K. Nagaraja, M. Rodehorst, S. Panchapagesan, N. Strom, S. Matsoukas, y S. Vitaladevuni. 2017. Compressed time delay neural network for small-footprint keyword spotting. En *INTERSPEECH*, páginas 3607–3611.

Zalman, M., L. L. Rubino, y B. Smith. 2019. Beyond police compliance with electronic recording of interrogation legislation: Toward error reduction. *Criminal Justice Policy Review*, 30(4):627–655.

Zhang, Y., G. Chen, D. Yu, K. Yaco, S. Khudanpur, y J. Glass. 2016. Highway long short-term memory rnns for distant speech recognition. En *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, páginas 5755–5759. IEEE.